Decision Making in Robots and Autonomous Agents

Decision Theory: Foundational Issues

Subramanian Ramamoorthy
School of Informatics

13 January, 2015
What is a Robot?

Problem: How to generate actions, to achieve high-level goals, using limited perception and incomplete knowledge of environment & adversarial actions?
Natural Language Dialogue as Closed Loop Decisions

System dynamically decides its actions

[Source: http://www.agilingua.com/images/DM_EN.jpg]
‘Robots’ in Virtual Worlds

[Schain+Mansour, AMEC 2012]
Computational Issues: Toy Example

Non-stationarity, plan recognition, personalisation, incentives, strategic coordination
Levels of Difficulty in Interaction

Consequences for *hardness of learning*:

1. Base case: spatial asymmetry
   - Learn a vector field

2. Next level: deal with reactive behaviour
   - ‘Inverse’ planning, plan recognition

3. Harder case: recursive exchange of beliefs (e.g., signaling, implicit coordination, trust, persuasion)
   - Need to model as a game?
In this course...

We will focus on how to **model** and **compute** decisions (*choices*),
- over **time**, under **uncertainty**, with **incompleteness** in models
- emphasizing **interactive** settings
- including methods for **learning** from experience and data.
- Also, we care about capturing how **real** people make choices!

Four Major Themes:
1. Decision/game theory and models
2. Learning
3. Preferences
4. Human choice behaviour
Course Structure

• Schedule of lectures is available at the course web site
  http://www.inf.ed.ac.uk/teaching/courses/dmr/
  I will attempt to upload slides the day before (except in first week)
• Two homework assignments
  – Pen-and-paper exercise on models, concepts, methods (20%)
  – Practical programming exercise in a mock-up domain (20%)
• Final Exam (60% of final mark)

• Resources:
  – No prescribed textbook
  – Some suggested readings (books) listed in course web site
  – Additional readings from books/articles for some lectures
Ask Questions!

– During the lecture
– After class, if your questions are brief
– After hours, by prior appointment only (arranged via email)
Teaser: Secretary Problem

• Choose one secretary from $n$ applicants
• Applicants interviewed sequentially in random order, each order being equally likely.
• Assume you can rank all applicants without ties. The decision to accept or reject must be based solely on the relative ranking of applicants interviewed so far. If you reach the final applicant, you are forced to hire that person.
• All decisions are instantaneous and final - an applicant already rejected can’t be reconsidered.
• Your decision criterion is to maximize the quality of the chosen candidate.
Solution to the Secretary Problem

• Wait until you have seen the first $n/e$ candidates and then pick the best one after that

• Why?
On what basis do you choose?

What does a theory need to account for?
Types of Decisions

• Who makes it?
  – Individual
  – ‘Group’

• What are the conditions?
  – Certainty
  – Risk
  – Uncertainty

For much of this course, we’ll take the ‘individual’ viewpoint (potentially acting in conflict of interest scenarios), and we’ll be somewhere between risk and uncertainty
How to Model Decision under *Certainty*?

- Given a set of possible acts
- Choose one that maximizes some given index

If a is a generic act in a set of feasible acts A, f(a) is an index being maximized, then

**Problem**: Find a* in A such that f(a*) > f(a) for all a in A.

The index f plays a key role, e.g., think of buying a painting.

Essential problem: How should the subject select an index function such that her choice reduces to finding maximizers?
An Operational Way to *Find* Index Function

• Observe subject’s behaviour in restricted settings and predict purchase behaviour from that:

• Instruct the subject as follows:
  – Here are ten valuable reproductions
  – We will present these to you in pairs
  – You will tell us which one of the pair you prefer to own
  – After you have evaluated all pairs, we will pick a pair at random and present you with the choice you previously made (it is to your advantage to remember your true tastes)

• The subject’s behaviour is as though there is a ranking over all paintings, so each painting can be summarized by a number
Some Properties of this Ranking

• *Transitivity*: Previous argument only makes sense if the rank is transitive – if A is preferred in (A, B) and B is preferred in (B, C) then A is preferred in (A, C); and this holds for all triples of alternatives A, B and C

• *Ordinal nature of index*: One is tempted to turn the ranking into a latent measure of ‘satisfaction’ but that is a mistake as utilities are non-unique.

  e.g., we could assign 3 utiles to A, 2 utiles to B and 1 utile to C to explain the choice behaviour

  Equally, 30, 20.24 and 3.14 would yield the same choice

While it is OK to compare indices, it is not OK to add or multiply
What Happens if we Relax Transitivity?

• Assume Pandora says (in the pairwise comparisons):
  – Apple < Orange
  – Orange < Fig
  – Fig < Apple

• Why is this a problem for Pandora?

• Assume a merchant who transacts with her as follows:
  – Pandora has an Apple at the start of the conversation
  – He offers to exchange Orange for Apple, if she gives him a penny
  – He then offers an exchange of Fig for Orange, at the price of a penny
  – Then, offers Apple for the Fig, for a penny
  – Now, what is Pandora’s net position?
Decision Making under *Risk*

- Initially appeared in mathematics as analysis of fair gambles, needed some notions of *utility*
- Gamble has $n$ outcomes, each worth $a_1, \ldots, a_n$
- The probability of each outcome is $p_1, \ldots, p_n$
- How much is it worth to participate in this gamble?
  \[ b = a_1 p_1 + \ldots + a_n p_n \]
  One may treat this monetary expected value as a fair price

Is this a sufficient description of choice behaviour under risk?
St. Petersburg Paradox of D. Bernoulli

• A fair coin is tossed until a head appears
• Gambler receives $2^n$ if the first head appears on trial $n$
• Probability of this event = probability of tail in first $(n-1)$ trials and head on trial $n$, i.e., $(1/2)^n$

Expected value $= 2 \cdot (1/2) + 4 \cdot (1/2)^2 + 8 \cdot (1/2)^3 + \ldots = \infty$

• Are you willing to bet in this way?
  – Is anyone?
Defining Utility

• Bernoulli went on to argue that people do not act in this way
• The thing to average is the ‘intrinsic worth’ of the monetary values, not the absolute values
  e.g., intrinsic worth of money may increase with money but at a *diminishing rate*

• Let us say utility of \( m \) is \( \log_{10} m \), then expected value is,

\[
\log_{10} 2 \cdot (1/2) + \log_{10} 4 \cdot (1/2)^2 + \log_{10} 8 \cdot (1/2)^3 + \ldots = b < \infty
\]

Monetary fair price of the gamble is \( a \) where \( \log_{10} a = b \).
Some Critiques of Bernoulli’s Formulation

von Neumann and Morgenstern (vNM), who seriously ‘started’ game theory, raised the following questions:

• The assignment of utility to money is arbitrary and *ad hoc*
  – There is an infinity of functions that capture ‘diminishing rate’, how should we choose?
  – The association may vary from person to person

• Why is the definition of the decision based upon *expected value* of the utility?
  – Is this actually descriptive of a single gambler, in an one-shot choice?
von Neumann & Morgenstern Formulation

• If a person is able to express preferences between every possible pair of gambles, where gambles are taken over some basic set of alternatives.

• Then one can introduce utility associations to the basic alternatives in such a manner that

• If the person is guided solely by the utility expected value, he is acting in accord with his true tastes.
  – provided his tastes are consistent in some way
Constructing Utility Functions

• Suppose we know the following preference order:
  – $a < b \sim c < d < e$

• The following are utility functions that capture this:

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>0</td>
<td>1/2</td>
<td>1/2</td>
<td>3/4</td>
<td>1</td>
</tr>
<tr>
<td>V</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>W</td>
<td>-8</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>8</td>
</tr>
</tbody>
</table>

– So, in situations like St Petersburg paradox, the revealed preference of any realistic player may differ from the case of infinite expected value
– Satisfaction at some large value, risk tolerance, time preference, etc.
Certainty Equivalents and Indifference

• The previous statement applies equally well to certain (‘deterministic’) events and gambles or lotteries
• So, even attitudes regarding tradeoffs between the two ought to be captured
• Basic issue – how to compare different “lotteries”?
• Imagine the following choice (A > B > C is preference order) : (a) you get B for certain, (b) A with prob. p and C otherwise
• If p is near 1, option b is better; if p is near 0, then option a: there is a single point where we switch
• Indifference is described as something like

\[
\frac{2}{3} (1) + \left(1 - \frac{2}{3}\right)(0) = \frac{2}{3}
\]
Caveats

• As before, we need to remember that the utility values should not be mis-interpreted
• The number 2/3 is determined by choices among risky alternatives and reflect attitude to ‘gambling’
• For instance, imagine a subject who would be indifferent to paying $9 and a 50-50 chance of paying $10 or nothing;
• This suggests utilities for $0, -$9, -$10 are 1, ½, 0.
• However, we can’t say it is just as enjoyable for him to go from -$10 to -$9 as it is to go from -$9 to $0!
• *Subject’s preferences among alternatives or lotteries come prior to numerical characterization of them*
Axiomatic Treatment of Utility

vNM and others formalize the above to define axioms for utility:

1) Any two alternatives shall be comparable, i.e., given any two, subject will prefer one over the other or be indifferent.

2) Both preference and indifference relations for lotteries are transitive.

3) In case a lottery has as one of its alternatives another lottery, then the first lottery is decomposable into the more basic alternatives through the use of the probability calculus.

4) If two lotteries are indifferent to the subject then they are interchangeable as alternatives in any compound lottery.
Axiomatic Treatment of Utility, contd.

vNM and others formalize the above to define axioms for utility:

5) If two lotteries involve the same two alternatives, then the one in which the more preferred alternative has a higher probability of occurring is itself preferred.

6) If A is preferred to B and B to C, then there exists a lottery involving A and C (with appropriate probabilities) which is indifferent to B.

… we’ll return to these issues later, especially when we discuss preference elicitation.
Decision Making under *Uncertainty*

- A choice must be made from among a set of acts, $A_1, \ldots, A_m$.
- The relative desirability of these acts depends on which state of nature prevails, either $s_1, \ldots, s_n$.
- As decision maker we know that one of several things is true and this influences our choice but we do not yet have a probabilistic characterization of these alternatives.

Savage’s omelet problem: Your friend has broken 5 good eggs into a bowl when you come in to volunteer and finish the omelet. A sixth egg lies unbroken (you must use it or waste it altogether). Your three acts: break it into bowl, break it into saucer – inspect and pour into bowl, throw it uninspected.
Decision Making under *Uncertainty*

To each outcome, we could assign a utility and maximize it.

What do we know about the state of nature?
- We may act *as though* there is one true state, we just don’t know it.
- If we assume a probability over $s$, this is decision under risk.

What criteria do we have for a decision problem under *uncertainty* (d.p.u.u.)?

---

![Table 1. Savage's example illustrating acts, states, and consequences](image.png)

<table>
<thead>
<tr>
<th>Act</th>
<th>Good</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Break into bowl</td>
<td>six-egg omelet</td>
<td>no omelet, and five good eggs destroyed</td>
</tr>
<tr>
<td>Break into saucer</td>
<td>six-egg omelet, and a saucer to wash</td>
<td>five-egg omelet, and a saucer to wash</td>
</tr>
<tr>
<td>Throw away</td>
<td>five-egg omelet, and one good egg destroyed</td>
<td>five-egg omelet</td>
</tr>
</tbody>
</table>
Risk vs. Uncertainty

Urn A Contains 100 Balls:

- 50 Red, 50 Black
- Pick A Color, Then Draw A Ball
- If You Draw Your Color, $10,000 Prize
- What Color Would You Prefer?
- How Much Would You Pay To Play?
Risk vs. Uncertainty

Urllib B Contains 100 Balls:
- Proportion Unknown
- Pick A Color, Then Draw A Ball
- If You Draw Your Color, $10,000 Prize
- What Color Would You Prefer?
- How Much Would You Pay To Play?

Knight’s (1921) Dichotomy of Risk vs. Uncertainty
A. Lo’s Extension of Knight’s Dichotomy

- Level 1: Complete Certainty
- Level 2: Risk without Uncertainty
- Level 3: Fully Reducible Uncertainty
- Level 4: Partially Reducible Uncertainty
- Level 5: Irreducible Uncertainty

Consider the Harmonic Oscillator

Simplest Non-Trivial Physical Model:

- Apply $F = ma$ to this relation:

\[
0 = \ddot{x} + \frac{k}{m} x
\]

\[x(t) = A \cos(\omega_o t + \phi), \quad \omega_o \equiv \sqrt{k/m}\]

Level 1: Certainty $\Rightarrow$

- At $t = 3.5$, we know $x = 1.7224$
Harmonic Oscillator

Level 2: Risk without Uncertainty \( \Rightarrow \)

\[ x(t) = A \cos(\omega_0 t + \phi) + \epsilon(t) \]

\( \epsilon(t) \text{ IID } \mathcal{N}(0, \sigma^2_\epsilon) \)

- At \( t = 3.5 \), we know
  \[ \text{Prob} \left( x \in [1.4284, 2.0164] \right) = 5\% \]
Harmonic Oscillator

Level 3: Fully Reducible Uncertainty

\[ x(t) = A \cos(\omega_0 t + \phi) + \epsilon(t) \]

\[ \mathbb{E}[x(t)] = 0 \]

\[ \mathbb{E}[\epsilon(t)\epsilon(s)] = \begin{cases} 
\sigma^2_{\epsilon} & \text{if } s \equiv t \\
0 & \text{otherwise}
\end{cases} \]

• Distribution of \( \epsilon(t) \) unknown but stationary and ergodic
Harmonic Oscillator

Level 4: Partially Reducible Uncertainty

- Two-state Markov-switching process
- Observer is unaware of the DGP

\[ x(t) = I(t) x_1(t) + (1 - I(t)) x_2(t) \]

\[ x_i(t) = A_i \cos(\omega_i t + \phi_i), \quad i = 1, 2 \]

\[ P \equiv \begin{pmatrix} I(t) = 1 & I(t) = 0 \\ I(t-1) = 1 & 1 - p & p \\ I(t-1) = 0 & p & 1 - p \end{pmatrix} \]
Regime Switching in Harmonic Oscillator
Regime Switching Dynamics

With additive noise

US GDP Data
Level 5: Irreducible Uncertainty (Unknowable)

- The “aliasing” or “identification” problem
- Many models may fit the same data, and no possibility of conducting controlled experiments
- This is a major factor in irreducible uncertainty
Applies to Fields of Knowledge:

1. Complete Certainty
2. Risk without Uncertainty
3. Fully Reducible Uncertainty
4. Partially Reducible Uncertainty
5. Irreducible Uncertainty

Mathematics
Physics
Chemistry
Biology
Economics
History
Philosophy
Religion
Some Criteria for d.p.u.u.

Maximin criterion: To each act, assign its security level as an index. Index of $A_i$ is the minimum of the utilities $u_{i1}, \ldots, u_{in}$

Choose the act whose associated index is maximum.

<table>
<thead>
<tr>
<th></th>
<th>s1</th>
<th>s2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>A2</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

- What is the security level for each act? Maximise minimum payoff.
- What happens if we allow for mixed strategies (i.e., akin to a compound lottery, e.g., $p = 0.5$ for $A1$ and $p = 0.5$ for $A2$)?
- Interpretation as game against nature: best response against nature’s minimax strategy (least favourable a priori strategy)
Point to Ponder about Maximin

• Is nature a *conscious* adversary?! 

• Consider:

<table>
<thead>
<tr>
<th></th>
<th>s1</th>
<th>s2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>A2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

• What are the safety values for the actions?  
  – If mixed strategies are allowed?

• What if 100 went up to $10^6$ and 1 came down to 0.0001?
Some Criteria for d.p.u.u.

- **Minimax risk criterion** (Savage): Consider a setup as follows:

<table>
<thead>
<tr>
<th></th>
<th>s1</th>
<th>s2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>A2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

If s1 is the true state, choosing A2 poses no ‘risk’ whereas if s2 is the true state then considerable ‘risk’ in A2.

Savage’s procedure: (i) Create new risk payoffs which are amounts to be added to utility to match maximum column utility, (ii) Choose act which minimizes maximum risk index.
Minimax Risk Criterion

• Transform Utility Payoff to Risk Payoff:

Take the $u_{ij}$ and define $r_{ij}$ so that it is the amount that has to be added to $u_{ij}$ to equal maximum utility payoff in column $j$.

• Critique (due to Chernoff):
  – “Regret” may not be measured by utility difference
  – Different states of nature may not be traded off properly
  – Taking away an irrelevant (obviously bad) action may change optimal decision!
More Criteria for d.p.u.u.

- **Pessimism-optimism index criterion** of Hurwicz:
  
  Let $m_i$ and $M_i$ be minimum and maximum utility. Assume a fixed pessimism-optimism index, $\alpha$. To each act, associate an $\alpha$-index $\alpha m_i + (1 - \alpha) M_i$.
  
  e.g., assign $\alpha$ based on indifference between actions, $a_i$.
  
  Of two acts, the one with higher $\alpha$-index is preferred.

- **“Principle of insufficient reason”**: If one is completely ignorant, one should act as though all states are equally likely; so choice should be based on a utility index which is the average of utility for all possible states for any act

*What is the effect of the way we enumerate possible states of nature?*
Acknowledgements

Much of the second half of this lecture is a paraphrased and condensed version of chapters in R.D. Luce and H. Raiffa’s excellent book, *Games and Decisions* (Dover Publishers, 1957; reprinted 1985)